

Epidemiological study designs

Causation

Epidemiology

- Descriptive - ecological
- Cross-sectional
- Cohort studies
 - Prospective
 - Retrospective
- Case-Control
- Experimental (intervention)

Epidemiological-Clinical

- Human subjects
- Clinical - focus on individual disease cases
- Experimental - controlled exposures/treatments
- Descriptive studies (simply describe)
 - Longitudinal (historical - follow over time)
 - Cross-sectional (snapshot in time)
- Analytical epidemiology (does a link exist?)
 - Cohort (start from exposure/treatment and look for disease)
 - Case-control (start with disease and compare exposures/treatments)
- Ecological or cluster investigations
 - Geographic, time period, certain population

Cohort studies

- Prospective (exposure now - disease to follow up in future)
- Retrospective (historical) (exposure in past - disease since then)
- Directly measure risk of a disease (illness rate, risk ratio or relative risk)
- Large groups needed (100's, 1000s) - costly
- ★ • Can assess many risk factors together
- Occupational more often
- Variant: proportional morbidity (cause of interest), best for uncommon diseases

Cohort study

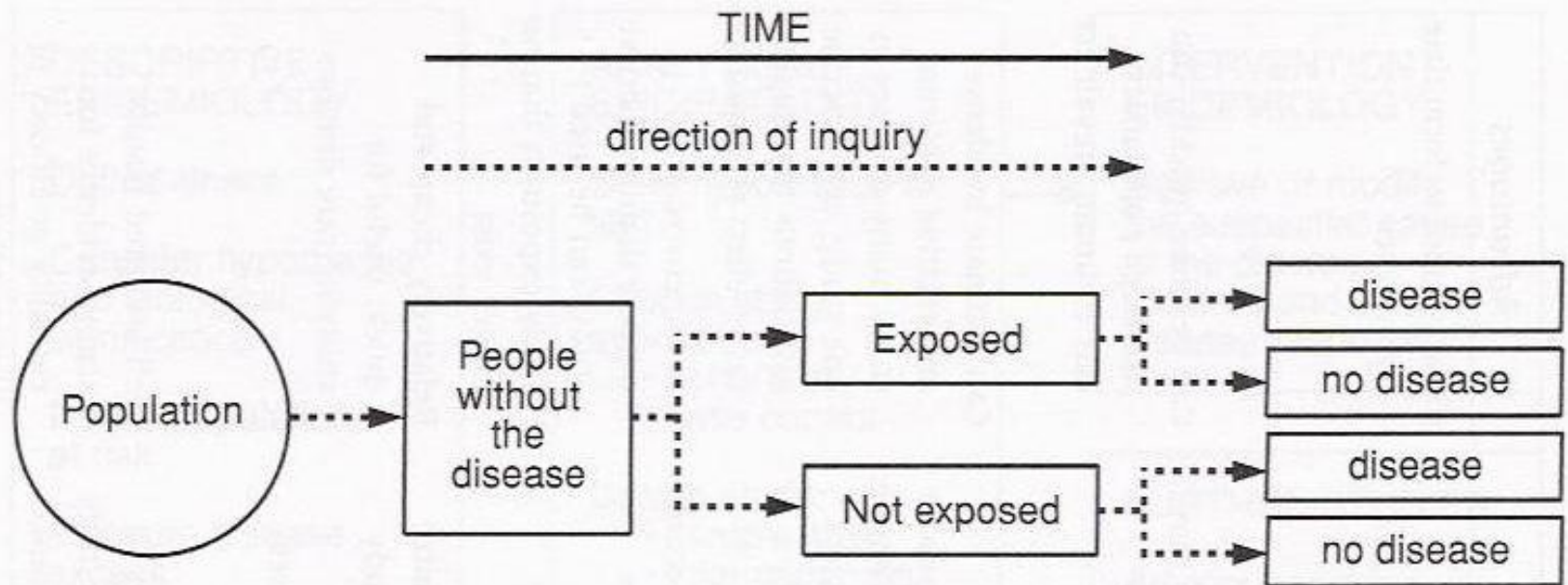


Figure 3.3 Design of a cohort study. From Beaglehole et al., 1993, with permission.

Case-control studies

- Powerful and accurate
- Economical (population size and time)
- Estimate odds ratios
- ★• Acute, chronic, long latency, rare diseases
- Variant: nested design (second phase, narrowed down)

Case-control study

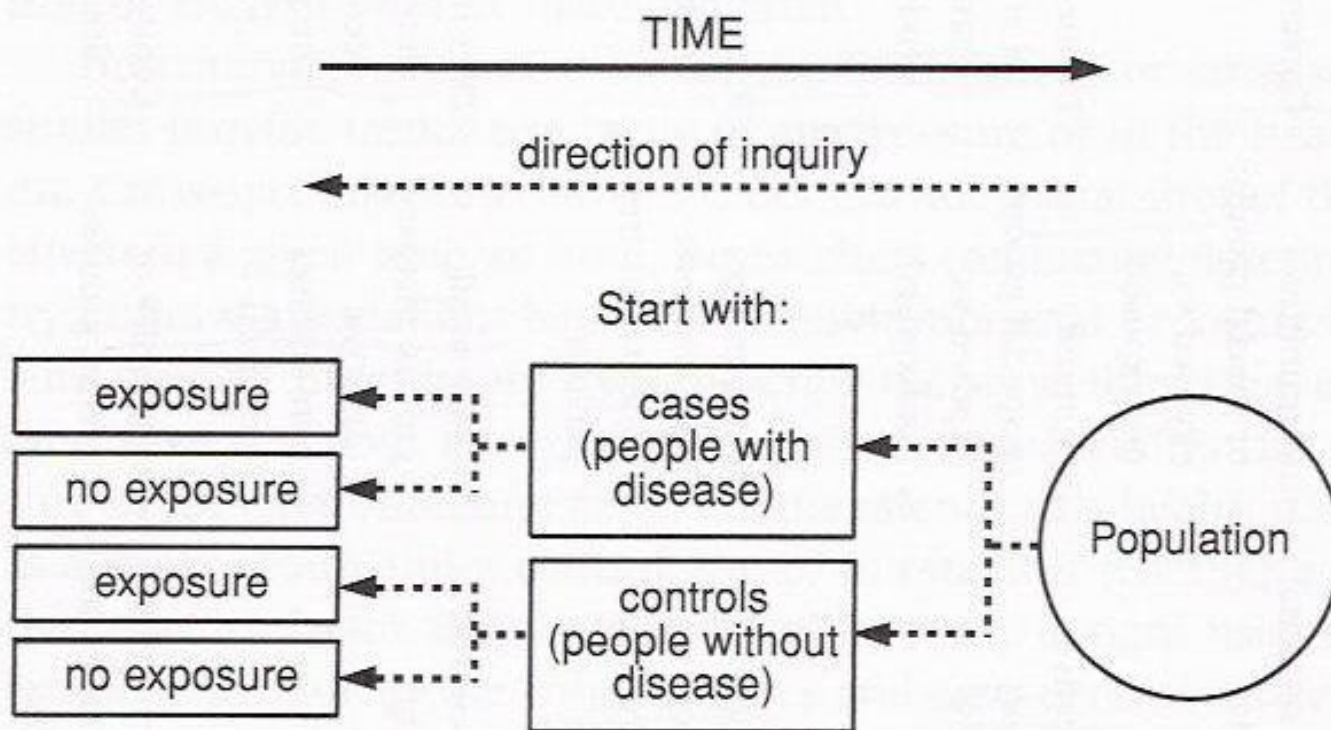


Figure 3.4 Design of a case-control study. From Beaglehole et al., 1993, with permission.

Ecological studies

- Unit of study is population group *not* individuals
- Most publicity
- Likely to be incorrect, misunderstood, or by chance
- Causation still necessary to prove, but ...
- Exposed and sick may ***not*** be the same individuals
- Useful in generating new hypotheses
- Inexpensive (existing data can be used)
- Exploratory, time-trend, space-time, multigroup, mixed

Rates of disease and disease ratios

RATE OF DISEASE:
(Incidence Rate)

$$\frac{\text{Number of cases of disease in population at risk}}{\text{Number of persons in population at risk}}$$

Expressed as:

$$\frac{\text{Number of Cases}}{100 \text{ or } 1000, 100,000 \text{ (usually) etc. persons at risk}}$$

Example:

$$\frac{50 \text{ Cases}}{2500 \text{ persons at risk}} = \frac{20}{1000}$$

RISK RATIO:
(RR = IR_E/IR_U)

$$\frac{\text{Rate of disease in population with the risk factor}}{\text{Rate of disease in population without the risk factor}} \\ \text{(comparison population)}$$

Expressed as:

A numerical ratio (1.5, 3.0 etc. indicating that risk of disease in the exposed (or at risk) population is 1.5, 3.0, etc. times greater than that in the unexposed (or not at risk) population

Example:

$$\frac{20 / 1000}{10 / 1000} = 2.0$$

Figure 3.5 Definition and calculation of rates of disease and risk ratios.

Measures

- Rate of disease: $\frac{\# \text{ cases}}{\# \text{ total at risk}}$
- Risk ratio: $\frac{\text{rate in exposed}^* (E)}{\text{rate in unexposed (U)}}$
- Attributable fraction: $\frac{E - U}{E} \text{ or } \frac{I - U}{I}$
- Odds ratio*: $\frac{A/C}{B/D}$

* Used in case-control studies

* treated/untreated

Mortality measure

- Standardized mortality ratios (SMR)

$$\frac{\text{Observed deaths}}{\text{Expected deaths}} \times 100$$

Causation

- Not simply an association in numbers but the plausibility that risk factor leads to disease (toxicological basis)
- Lab studies required to supplement epidemiological studies
- Large enough study
(2x increase, $n=300$, for $\alpha=5\%$, $\beta=20\%$ or 80% chance to detect a true effect)

Does association mean the exposure caused the effect?

TABLE 3.3

TESTS OF CAUSATION

Temporal relation: Does the cause precede the effect? (essential)

Plausibility: Is the association consistent with other knowledge?

Mechanism of action: Is there evidence from experimental animals?

Consistency: Have similar results been shown in other studies?

Strength: What is the strength of the association between the cause and the effect? (relative risk)

Dose-response relationship: Is increased exposure to the possible cause associated with increased effect?

Reversibility: Does the removal of a possible cause lead to reduction of disease risk?

Study design: Is the evidence based on a strong study design?

Judging the evidence: How many lines of evidence lead to the conclusion?

Source: Beaglehole et al., 1993 (these are modified criteria of causation from those originally developed by Bradford Hill).

Statistical concepts

- Central tendency (average value or mean, median)
- Accuracy (bounds of likely values, ie. within +/- x%)
 - Confidence interval (likely range of values)
 - Depends on # cases, population size, variability,
- Errors
 - Alpha (α): observing the effect by chance alone (see an effect when there is none)
 - Beta (β): not being able to detect a true effect (lack of sensitivity in detection of effect)
- Power: being able to detect a true effect ($1-\beta$) (size of study and variability the most influential along with rarity of outcome)

How many people do we need so that we can detect a doubling of incidence?

TABLE 3.4

SAMPLE SIZE REQUIRED TO DETECT A DOUBLING OF BACKGROUND
INCIDENCE IN REPRODUCTIVE OUTCOME

<i>Reproductive Outcome</i>	<i>Size of Each Group Required^a</i>
Infertility	161 couples
Spontaneous abortion	161 pregnancies
Stillbirth	161 pregnancies
Low birth weight	293 live births
Major birth defects	316 live births
Infant deaths	928 live births
Severe mental retardation	4493 live births
Chromosome abnormalities	8951 live births

^aWith $\alpha = 0.05$; $\beta = 0.20$.

Source: NIOSH, 1988.

TABLE 3.2

STUDY DESIGNS IN ENVIRONMENTAL EPIDEMIOLOGY THAT USE THE INDIVIDUAL AS THE UNIT OF ANALYSIS

<i>Study Design</i>	<i>Population</i>	<i>Exposure</i>	<i>Health Effect</i>	<i>Confounders</i>	<i>Problems</i>	<i>Advantages</i>
Descriptive study	Community or various subpopulations	Records of past measurements	Mortality and morbidity statistics; case registries; other reports	Difficult to sort out	Difficult to establish exposure-effect relationships	Cheap, useful to formulate hypotheses
Cross-sectional study	Communities or special groups; exposed vs. nonexposed	Current	Current	Usually easy to measure	Current exposure may be irrelevant to current disease	Can be done quickly; can use large populations; <i>can estimate prevalence</i>
Prospective cohort study	Community or special groups; exposed vs. nonexposed	Defined at outset of study (can change during study)	To be determined during study	Usually easy to measure	Expensive, time consuming; exposure categories can change; high dropout rate possible	Can estimate incidence and relative risk; can study many diseases in one study; can describe associations that suggest cause-effect relationships
Historical cohort study	Special groups, e.g., workers, patients, insured persons	Records of past measurement	Records of past or current diagnosis	Often difficult to measure because of retrospective nature; depends on quality of previously obtained data	Need to rely on records that may not be accurate	Less expensive and quicker than prospective study; can be used to study exposures that no longer exist
Case-control study	Usually small groups; diseases (cases) vs. non-diseases (controls)	Occurred in past, determined by records or interview	Known at start of study	Possible to eliminate by matching for them	Difficult to generalize due to small study groups; some incorporate biases	Relatively cheap and quick; useful for studying rare diseases
Experimental (intervention study)	Community or special groups	Controlled/known already	To be measured during study	Can be controlled by randomization of subjects	Expensive; ethical considerations; study subjects' compliance required	Well-accepted results; strong evidence for causality or efficacy of intervention